

28.10.2004

日本国特許庁
JAPAN PATENT OFFICE

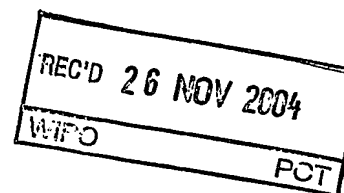
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年10月29日
Date of Application:

出願番号 特願2003-368773
Application Number:
[ST. 10/C]: [JP 2003-368773]

出願人 日本アイ・ビー・エム株式会社
Applicant(s):

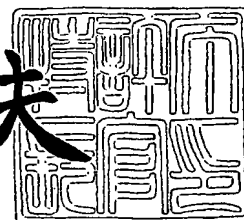


**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1 (a) OR (b)

2004年 6月 7日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



【書類名】 特許願
【整理番号】 JP9030230
【提出日】 平成15年10月29日
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 15/177
【発明者】
 【住所又は居所】 東京都港区六本木3丁目2番12号 日本アイ・ビー・エム株式会社内
 【氏名】 井上 正志
【特許出願人】
 【識別番号】 592073101
 【氏名又は名称】 日本アイ・ビー・エム株式会社
【代理人】
 【識別番号】 100086243
 【弁理士】
 【氏名又は名称】 坂口 博
【代理人】
 【識別番号】 100091568
 【弁理士】
 【氏名又は名称】 市位 嘉宏
【代理人】
 【識別番号】 100108501
 【弁理士】
 【氏名又は名称】 上野 剛史
【復代理人】
 【識別番号】 100104156
 【弁理士】
 【氏名又は名称】 龍華 明裕
【手数料の表示】
 【予納台帳番号】 053394
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1
 【包括委任状番号】 9706050
 【包括委任状番号】 9704733
 【包括委任状番号】 0207860

【書類名】 特許請求の範囲**【請求項 1】**

端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システムであって、

前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間を監視する処理時間監視部と、

前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあると特定するボトルネック特定部と、

ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる負荷制御部とを備える情報システム。

【請求項 2】

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記アプリケーションサーバ数未満のしきい値以下の数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でないことを条件として、当該アプリケーションサーバにボトルネックがあると特定し、前記しきい値を超える数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でない場合に、前記データベースサーバにボトルネックがあると特定し、

前記負荷制御部は、当該アプリケーションサーバにボトルネックがあると特定された場合に、当該アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させ、前記データベースサーバにボトルネックがあると特定された場合に、前記複数のアプリケーションサーバのそれぞれについて前記アプリケーションプログラムを実行する多重度を低下させる

請求項 1 記載の情報システム。

【請求項 3】

前記複数のアプリケーションサーバのそれぞれは、当該アプリケーションサーバ上で実行されるべき優先度が設定された複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、

前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でないことを条件として、当該アプリケーションサーバにボトルネックがあると特定し、

前記負荷制御部は、ボトルネックがあると特定された前記アプリケーションサーバ上で前記一の前記アプリケーションプログラムより優先度が低い前記アプリケーションプログラムを実行する多重度を低下させる

請求項 1 記載の情報システム。

【請求項 4】

前記複数のアプリケーションサーバのそれぞれは、当該アプリケーションサーバ上で実行されるべき優先度が設定された複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、

前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、予め定められたしきい値を超える数の前記アプリケーションサーバについて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でない場合に、前記データベースサーバにボトルネックがあると特定し、

前記負荷制御部は、ボトルネックがあると特定された前記データベースサーバにおける、前記一のアプリケーションプログラムより優先度が低い前記アプリケーションプログラムの要求に基づき前記データベースをアクセスする処理の優先度を低下させる

請求項1記載の情報システム。

【請求項5】

前記複数のアプリケーションサーバのそれぞれは、当該アプリケーションサーバ上で実行されるべき優先度が設定された複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、

前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、予め定められたしきい値を超える数の前記アプリケーションサーバについて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でない場合に、前記データベースサーバにボトルネックがあると特定し、

前記負荷制御部は、ボトルネックがあると特定された前記データベースサーバにおける、前記一のアプリケーションプログラムからの要求に基づき前記データベースをアクセスする処理の優先度を低下させる

請求項1記載の情報システム。

【請求項6】

前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが有する、前記アプリケーションプログラムにより使用される資源であるアプリケーションサーバ資源の使用量を監視する処理負荷監視部を更に備え、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記アプリケーションサーバ数未満のしきい値以下の数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でない場合に、当該アプリケーションサーバにボトルネックがあると特定し、

前記負荷制御部は、ボトルネックがあると特定された前記アプリケーションサーバにおける前記アプリケーションサーバ資源の使用量が予め定められた許容範囲内でないことを条件として、ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる

請求項1記載の情報システム。

【請求項7】

前記処理負荷監視部は、前記データベースサーバが有する、前記アプリケーションプログラムが前記データベースをアクセスする場合に使用される資源であるデータベースサーバ資源の使用量を更に監視し、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記しきい値を超える数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でないことを条件として、前記データベースサーバにボトルネックがあると特定し、

前記負荷制御部は、ボトルネックがあると特定された前記データベースサーバにおける前記データベースサーバ資源の使用量が予め定められた許容範囲内でない場合に、前記複数のアプリケーションサーバのそれぞれについて当該アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる

請求項 6 記載の情報システム。

【請求項 8】

前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが有する、前記アプリケーションプログラムにより使用される資源である複数のアプリケーションサーバ資源の使用量と、前記データベースサーバが有する、前記アプリケーションプログラムが前記データベースをアクセスする場合に使用される資源である複数のデータベースサーバ資源の使用量とを監視する処理負荷監視部を更に備え、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記複数のアプリケーションサーバ及び前記データベースサーバのうち少なくとも 1 つの前記アプリケーションサーバ又は前記データベースサーバにボトルネックがあると特定し、

前記負荷制御部は、

前記アプリケーションサーバにおける前記複数のアプリケーションサーバ資源、及び、前記データベースサーバにおける前記複数のデータベースサーバ資源のそれぞれに対応して、当該資源を有するサーバにボトルネックがあると特定され、かつ、当該資源の使用量が予め定められた許容範囲内でない場合に前記負荷制御部が行うべき負荷制御処理を指定する負荷制御情報を格納する負荷制御情報格納部と、

ボトルネックがあると特定されたサーバが有する資源の使用量が予め定められた許容範囲内でない場合に、当該資源がボトルネックであると特定し、前記アプリケーションサーバ資源がボトルネックであると特定された場合に、当該資源に対応して前記負荷制御情報格納部に格納された前記負荷制御情報を取得し、前記データベースサーバ資源がボトルネックであると特定された場合に、当該資源に対応して前記負荷制御情報格納部に格納された前記負荷制御情報を取得する負荷制御情報取得部と、

ボトルネックがあると特定されたサーバのボトルネックを解消するべく、前記負荷制御情報により指定される前記負荷制御処理を実行する負荷制御処理部と

を有する

請求項 1 記載の情報システム。

【請求項 9】

前記複数のアプリケーションサーバのそれぞれは、複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、

前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、

前記処理負荷監視部は、前記データベースサーバについて、当該データベースサーバが有する前記データベースサーバ資源であるプロセッサ及びデータベースの格納されているディスクの使用量を監視し、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、前記しきい値を超える数の前記アプリケーションサーバについて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でないことを条件として、前記データベースサーバにボトルネックがあると特定し、

前記負荷制御情報格納部は、前記データベースサーバにおける前記アプリケーション資源であるプロセッサに対応する第 1 の前記負荷制御情報と、データベースの格納されているディスクに対応する第 2 の前記負荷制御情報とを格納し、

前記負荷制御情報取得部は、ボトルネックがあると特定された前記データベースサーバが有するプロセッサの使用量が予め定められた使用量を超える場合に前記第 1 の負荷制御情報を取得し、ボトルネックがあると特定された前記データベースサーバが有するデータベースの格納されているディスクの使用量が予め定められた使用量を超える場合に前記第 2 の負荷制御情報を取得し、

前記負荷制御処理部は、前記負荷制御情報取得部から前記第 1 の負荷制御情報が取得さ

れた場合に、前記複数のアプリケーションサーバのそれぞれについて前記一のアプリケーションプログラムを実行する多重度を低下させ、前記負荷制御情報取得部から前記第2の負荷制御情報が取得された場合に、前記データベースサーバにおいて当該制御情報で指定された当該ディスクを使用するアプリケーションプログラムのグループの中で、前記一のアプリケーションプログラムより優先度が低い前記アプリケーションプログラムからの要求に基づき前記データベースをアクセスする処理の優先度を低下させる

請求項8記載の情報システム。

【請求項10】

前記負荷制御情報格納部に格納された一の前記負荷制御情報は、複数の前記負荷制御処理と前記複数の負荷制御処理の優先順位とを指定する情報であり、

前記負荷制御処理部は、前記一の負荷制御情報により指定された一の前記負荷制御処理を実行した後、ボトルネックがあると特定されたサーバのボトルネックが予め設定された設定期間の間に解消しないことを検出した場合に、前記一の負荷制御情報より指定される前記一の負荷制御処理より優先順位が低い他の前記負荷制御処理を実行する

請求項8記載の情報システム。

【請求項11】

前記複数のアプリケーションサーバのそれぞれは、複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、

前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、

前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、前記アプリケーションサーバ数未満のしきい値以下の前記アプリケーションサーバについて一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でない場合に、当該アプリケーションサーバにボトルネックがあると特定し、

前記負荷制御部は、ボトルネックがあると特定された前記アプリケーションサーバ上で前記一のアプリケーションプログラムより低い優先度に設定された前記アプリケーションプログラムを実行する多重度を低下させ、当該アプリケーションサーバのボトルネックが前記設定期間の間に解消しないことを検出した場合に、当該アプリケーションサーバ上で前記一のアプリケーションプログラムを実行する多重度を更に低下させる

請求項8記載の情報システム。

【請求項12】

端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システムをコンピュータにより制御する負荷制御方法であって、

前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間をコンピュータにより監視する処理時間監視段階と、

前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあるとコンピュータにより特定するボトルネック特定段階と、

ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度をコンピュータにより低下させる負荷制御段階と

を備える負荷制御方法。

【請求項13】

端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システム

をコンピュータにより制御させる負荷制御プログラムであって、

前記コンピュータを、

前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間を監視する処理時間監視部と、

前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあると特定するボトルネック特定部と、

ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる負荷制御部と

して機能させる負荷制御プログラム。

【請求項 14】

請求項 13 に記載の負荷制御プログラムを記録した記録媒体。

【書類名】明細書

【発明の名称】情報システム、負荷制御方法、負荷制御プログラム及び記録媒体

【技術分野】

【0001】

本発明は、情報システム、負荷制御方法、負荷制御プログラム及び記録媒体に関する。特に本発明は、複数のアプリケーションサーバ及びデータベースサーバを備え、負荷制御を行う情報システムと、その負荷制御方法、負荷制御プログラム及び記録媒体とに関する。

【背景技術】

【0002】

近年、Webページの表示と連携してサーバサイドでアプリケーションを実行可能とする各種の基盤技術が提供されていることから、利用者に良好なユーザインターフェイスを提供するWeb機能を用いて情報システムを実現することが主流となっている。

このような情報システムの負荷を制御する技術として、1台のコンピュータ上でサーバプログラムを実行させ、サーバプログラムが当該コンピュータのCPUを使用する使用率が高い場合にサーバプログラムの多重度を増加させる技術が開示されている（特許文献1）。

また、LAN接続された負荷分散装置とサーバコンピュータとを備えるクライアントサーバシステムにおいて、サーバコンピュータの負荷がしきい値を超える場合に、負荷分散装置が端末からのトランザクションの受け入れを中断する技術が開示されている（特許文献2）。

【特許文献1】特開2001-160040号公報

【特許文献2】特開2000-268012号公報

【発明の開示】

【発明が解決しようとする課題】

【0003】

大規模な情報システムは、複数のアプリケーションサーバと、1台又は多重化されたデータベースサーバとを備えて構成される。アプリケーションサーバは、利用者によりWebページがアクセスされた場合に当該Webページに対応付けられたアプリケーションを実行し、実行結果を含めたWebページのデータを当該端末へ返信する。データベースサーバは、アプリケーションサーバからデータベースのアクセス要求を受けてデータベースをアクセスする。このような情報システムにおいては、利用者の端末から受信したトランザクションに対するレスポンス時間を許容範囲内に抑え、利用者の作業に遅延を生じさせないことが重要である。しかしながら、上記の技術においては、サーバプログラムやサーバコンピュータの負荷を一定以下に保つことを目的としており、トランザクションのレスポンス時間に基づいた制御を行うことができず、また、アプリケーションサーバ及びデータベースサーバのそれぞれに適した負荷制御処理を行うことができなかった。

【0004】

そこで本発明は、上記の課題を解決することのできる情報システム、負荷制御方法、負荷制御プログラム及び記録媒体を提供することを目的とする。この目的は特許請求の範囲における独立項に記載の特徴の組み合わせにより達成される。また従属項は本発明の更なる有利な具体例を規定する。

【課題を解決するための手段】

【0005】

本発明の第1の形態によると、端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システムであって、前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間を監視する処理時間監視部と、前記複数のアプリ

ケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあると特定するボトルネック特定部と、ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる負荷制御部とを備える情報システムと、当該情報システムに関する負荷制御方法、負荷制御プログラム、及び記録媒体を提供する。

【0006】

なお、上記の発明の概要は、本発明の必要な特徴の全てを列挙したものではなく、これらの特徴群のサブコンビネーションもまた、発明となりうる。

【発明の効果】

【0007】

本発明によれば、複数のアプリケーションサーバ及びデータベースサーバを備える情報システムにおいて、各アプリケーションサーバにおけるトランザクションの処理時間に応じて適切な負荷制御を行うことができる。

【発明を実施するための最良の形態】

【0008】

以下、発明の実施の形態を通じて本発明を説明するが、以下の実施形態は特許請求の範囲にかかる発明を限定するものではなく、また実施形態の中で説明されている特徴の組み合わせの全てが発明の解決手段に必須であるとは限らない。

【0009】

図1は、本実施形態に係る情報システム10の構成を示す。本実施形態に係る情報システム10は、複数のアプリケーションサーバ120のそれぞれにおけるトランザクションの処理時間と、複数のアプリケーションサーバ120及びデータベースサーバ150の処理負荷とを監視し、監視結果に基づいてアプリケーションサーバ120又はデータベースサーバ150のボトルネックを解消することができる適切な負荷制御を行う。また、本実施形態に係る情報システム10は、複数種類の業務等に対応する複数種類のアプリケーションプログラムを統合して稼動させた場合において、あるアプリケーションプログラムに生じた性能低下要因が他のアプリケーションプログラムの処理性能に影響を与えないように適切に負荷制御を行うことができる。

【0010】

情報システム10は、当該情報システム10を利用する利用者がそれぞれ使用する複数の端末100にインターネットやイントラネット等のネットワーク110を介して接続されるサーバコンピュータシステムであり、複数のアプリケーションサーバ(WAS: Web Application Server)120、データベースサーバ(DBS: Database Server)150、及び負荷制御サーバ160を備える。

複数のWAS120のそれぞれは、端末100から受信したHTTPリクエスト等のトランザクションに応じて、Java(登録商標)サーブレットやJSP(Java Server Pages)等によるアプリケーションプログラムを実行するサーバコンピュータである。また、WAS120は、実行結果を含めたWebページを端末100へ返信する。このアプリケーションプログラムの実行中に、WAS120は、アプリケーションプログラムによる指示に基づきデータベース(DB)152のアクセスをDBS150に要求する。

DBS150は、複数のWAS120のそれぞれの要求に基づきDB152をアクセスし、アクセス結果を要求元のWAS120に返信するサーバコンピュータである。負荷制御サーバ160は、複数のWAS120のそれぞれにおけるトランザクションの処理時間と、複数のWAS120及びDBS150の処理負荷とを監視し、監視結果に基づいて情報システム10の負荷制御を行う。

【0011】

次に、WAS120、DBS150、及び負荷制御サーバ160の構成を示す。

複数のWAS120のそれぞれは、HTTPサーバ部125と、WEBアプリケーションサーバ部130と、処理時間取得部140と、処理負荷取得部142と、サーバ内負荷

制御部144とを有する。HTTPサーバ部125は、端末100からトランザクションを受信し、当該トランザクションに対応するWebページを端末100へ返信する。なお、当構成においてHTTPサーバ部125は、物理的に別の筐体であるサーバ上で稼働している。

WEBアプリケーションサーバ部130は、端末100から受信したトランザクションに応じてアプリケーションプログラム134を実行する。本実施形態において、複数のWAS120のそれぞれにおけるWEBアプリケーションサーバ部130は、複数種類のアプリケーションプログラム134を混載し、複数種類のアプリケーションプログラム134のうち、端末100から受信したトランザクションの種類に対応するアプリケーションプログラム134を実行する。ここで複数種類のアプリケーションプログラム134は、当該WAS120上で実行されるべき優先度がそれぞれ設定される。より具体的には、複数種類のアプリケーションプログラム134は、優先度に応じて複数のグループに分類され、各アプリケーションプログラム134が属するグループに指定された優先度で実行される。

【0012】

WEBアプリケーションサーバ部130は、多重度制御部132、DBアクセス制御部136、及びDBコネクションプール138を含む。多重度制御部132は、複数種類のアプリケーションプログラム134のそれぞれを実行すべき多重度を制御する。ここでアプリケーションプログラム134を実行する多重度とは、同一種類のトランザクションを複数受信した場合に、それぞれのトランザクションに対応してアプリケーションプログラム134の並行して動作させる数の最大値である。

DBアクセス制御部136は、複数種類のアプリケーションプログラム134のそれぞれによるDB152のアクセス要求を併せて処理する。そしてDBアクセス制御部136は、DBS150に対するアクセス要求の発行数をアプリケーションプログラム134毎に制限すると共に、DBS150内部で複数種類のアプリケーションプログラム134のそれぞれの優先度を特定可能とするようにDBコネクション139を割り当てる。より具体的には、DBアクセス制御部136は、優先度毎に異なるユーザIDを用いてDBS150に接続することにより、DBS150内部でアクセス要求の優先度を特定可能とする。

DBコネクションプール138は、複数のDBコネクション139を含む。それぞれのDBコネクション139は、DB152のアクセス要求を発行するためにDBS150と接続する接続点となる。

以上に示した多重度制御部132及びDBアクセス制御部136は、1台のWAS120上で複数種類のアプリケーションプログラム134を統合して効率良く稼働させるためのフレームワークを実現するクラス関数等を、WAS120上で実行することにより実現されてよい。

【0013】

処理時間取得部140は、個々のトランザクションの処理時間やデータベース処理時間等を多重度制御部132およびDBアクセス制御部136から取得し、アプリケーションプログラム134毎の処理時間及びデータベース処理時間等の統計を取る。ここで、端末100から受信したトランザクションをアプリケーションプログラム134が処理する処理時間は、トランザクションのレスポンス時間とも呼ばれる。そして、処理時間取得部140は、一のアプリケーションプログラム134の処理時間が許容範囲内でない場合に、当該アプリケーションプログラム134についてサービスレベルが低下していることを示すサービスレベル低下イベントを負荷制御サーバ160へ通知する。ここで処理時間取得部140は、このサービスレベル低下イベントにデータベース処理時間等の情報を含めて負荷制御サーバ160へ通知してよい。

【0014】

処理負荷取得部142は、当該WAS120が有する、アプリケーションプログラム134によりそれぞれ使用される資源である1又は複数のアプリケーションサーバ資源の使

用量を、例えばWAS120上で動作するオペレーティングシステムのパフォーマンスモニタ機能等から取得し、当該アプリケーションサーバ資源の使用量の統計を取る。そして、処理負荷取得部142は、一のアプリケーションサーバ資源の使用量が許容範囲内で無い場合に、当該アプリケーションサーバ資源の処理負荷が高負荷であることを示すシステム資源高負荷イベントを負荷制御サーバ160へ通知する。以上において処理負荷取得部142は、アプリケーションサーバ資源の使用量をアプリケーションプログラム134毎に取得し、当該アプリケーションプログラム134による使用量が許容範囲で無い場合にシステム資源高負荷イベントを負荷制御サーバ160へ通知してもよく、これに代えて処理負荷取得部142は、全てのアプリケーションプログラム134によるアプリケーションサーバ資源の使用量の合計を取得し、当該使用量が許容範囲内で無い場合にシステム資源高負荷イベントを負荷制御サーバ160へ通知してもよい。

【0015】

サーバ内負荷制御部144は、負荷制御サーバ160により実行される負荷制御処理に応じて、当該WAS120内における負荷制御を行う。本実施形態に係るサーバ内負荷制御部144は、WAS120内における負荷制御として、各アプリケーションプログラム134のそれぞれを実行する多重度を変更する。

【0016】

DBS150は、DB152と、DB処理要求I/F部153と、DBエンジン154と、処理負荷取得部156と、サーバ内負荷制御部158とを有する。DB152は、複数のアプリケーションプログラム134のそれぞれによりアクセスされるデータを格納する。複数のDB処理要求I/F部153のそれぞれは、複数のDBコネクション139のそれぞれに対応して設けられ、WAS120がDB152に対するアクセス要求を発行するための接続点となる。本実施形態において、WAS120は、複数のアプリケーションプログラム134のそれぞれの優先度に応じて異なるユーザIDに対応するDB処理要求I/F部153に接続する。このためDBS150は、ユーザIDに基づいてアクセス要求を発行したアプリケーションプログラム134の優先度を識別することができる。

【0017】

DBエンジン154は、複数のWAS120のそれぞれが複数のアプリケーションプログラム134のそれぞれの指示に基づき発行するSQLコマンド等のアクセス要求に基づき、DB152をアクセスする。処理負荷取得部156は、DBS150が有する、アプリケーションプログラム134がDB152をアクセスする場合に使用される資源である1又は複数のデータベースサーバ資源の使用量を、例えばDBS150上で動作するオペレーティングシステムのパフォーマンスモニタ機能から取得し、当該データベースサーバ資源の使用量の統計を取る。そして、処理負荷取得部156は、一のデータベースサーバ資源の使用量が許容範囲内でない場合に、当該データベースサーバ資源の処理負荷が高負荷であることを示すシステム資源高負荷イベントを負荷制御サーバ160へ通知する。以上において処理負荷取得部156は、データベースサーバ資源の使用量をDB処理要求I/F部153毎に取得し、対応するアプリケーションプログラム134による使用量が許容範囲内でない場合にシステム資源高負荷イベントを負荷制御サーバ160へ通知してもよく、これに代えて処理負荷取得部156は、全てのDB処理要求I/F部153によるデータベースサーバ資源の使用量の合計を取得し、当該使用量が許容範囲内でない場合にシステム資源高負荷イベントを負荷制御サーバ160へ通知してもよい。

【0018】

サーバ内負荷制御部158は、負荷制御サーバ160により実行される負荷制御処理に応じて、DBS150内における負荷制御を行う。本実施形態に係るサーバ内負荷制御部158は、サーバ内負荷制御部158内における負荷制御として、DB処理要求I/F部153がDB152をアクセスする処理の優先度を変更する。

【0019】

負荷制御サーバ160は、処理時間監視部165と、ボトルネック特定部170と、処理負荷監視部190と、負荷制御部175とを有する。処理時間監視部165は、複数の

WAS 120のそれぞれについて、当該WAS 120が受信したトランザクションをアプリケーションプログラム134が処理する処理時間を監視する。本実施形態において、処理時間監視部165は、複数のWAS 120のそれぞれについて、アプリケーションプログラム134毎のトランザクションの処理時間を監視する。ここで処理時間監視部165は、複数のWAS 120のそれぞれから通知されるサービスレベル低下イベントに基づいて、アプリケーションプログラム134毎のトランザクションの処理時間が許容範囲外となったことを検出する。

【0020】

ボトルネック特定部170は、複数のWAS 120のそれぞれについての処理時間の監視結果に基づいて、複数のWAS 120及びDBS 150のうち少なくとも1つのWAS 120又はDBS 150にボトルネックがあると特定する。処理負荷監視部190は、複数のWAS 120のそれぞれについての1又は複数のアプリケーションサーバ資源の使用量と、1又は複数のデータベースサーバ資源の使用量とを監視する。本実施形態において、処理負荷監視部190は、複数のWAS 120及びDBS 150のそれぞれから通知されるシステム資源高負荷イベントに基づいて、当該システム資源高負荷イベントにより特定されるアプリケーションサーバ資源又はデータベースサーバ資源の使用量が許容範囲外となったことを検出する。

【0021】

負荷制御部175は、複数のWAS 120又はDBS 150の少なくとも1つのサーバにボトルネックがあると特定された場合に、ボトルネックがあると特定されたサーバのボトルネックを解消するための負荷制御処理を実行する。

負荷制御部175は、負荷制御情報格納部177と、負荷制御情報取得部179と、負荷制御処理部181とを含む。負荷制御情報格納部177は、ボトルネックと特定されたサーバと、当該サーバにおいて高負荷であることを検出された資源とに対応して、負荷制御部175内の負荷制御処理部181が行うべき負荷制御処理を指定する負荷制御情報を格納する。さらに、負荷制御情報格納部177は、アプリケーションプログラム134のそれぞれのグループ、グループの優先度、及び、当該アプリケーションプログラム134に設定可能な多重度を示す多重度情報を格納する。負荷制御情報取得部179は、ボトルネック特定部170によりボトルネックと特定されたサーバにおける、処理負荷監視部190により高負荷であることが検出された資源に対応する負荷制御情報を負荷制御情報格納部177から取得し、負荷制御処理部181へ出力する。負荷制御処理部181は、ボトルネックがあると特定されたサーバのボトルネックを解消するべく、負荷制御情報により指定される負荷制御処理を実行する。負荷制御処理において、負荷制御処理部181は、負荷制御の対象となるWAS 120内のサーバ内負荷制御部144又はDBS 150内のサーバ内負荷制御部158に対して、当該サーバ内の負荷制御を指示する。

【0022】

以上に示した情報システム10によれば、複数のアプリケーションサーバ120のそれぞれにおけるトランザクションの処理時間と、複数のアプリケーションサーバ120及びデータベースサーバ150の処理負荷とを監視し、監視結果に基づいてボトルネックを解消する負荷制御処理を適切に選択し実行することができる。

【0023】

図2は、本実施形態に係る情報システム10のボトルネック特定フローを示す。

まず、処理時間監視部165は、1又は複数のWAS 120から1又は複数のサービスレベル低下イベントを通知されることにより、トランザクションの処理時間が許容範囲外となったことを検出する。ボトルネック特定部170は、複数のWAS 120のそれぞれについての処理時間の監視結果に基づいて、ボトルネックの存在箇所を特定する。より具体的にはボトルネック特定部170は、複数のWAS 120のそれぞれについてのアプリケーションプログラム134毎の処理時間の監視結果に基づいて、ボトルネックの存在箇所を特定する。

【0024】

まず、ボトルネック特定部170は、1以上しきい値以下の数のWAS120について処理時間が許容範囲内でないことを条件として（ステップS200:Yes）、処理時間が許容範囲内でないWAS120にボトルネックがあると特定する（S220）。ここでボトルネック特定部170は、1以上しきい値以下の数のWAS120について一のアプリケーションプログラム134の処理時間が許容範囲内でない場合（S200:Yes）に、処理時間が許容範囲内でないWAS120における、他のアプリケーションプログラム134の処理時間も許容範囲内でないことを更に条件として（S210:Yes）当該WAS120にボトルネックがあると特定し（S220）、他のアプリケーションプログラム134の処理時間が許容範囲内であれば（S210:No）、当該アプリケーションプログラム134に対するトランザクションの特定の入力に関連する資源がボトルネックであると特定してもよい（S230）。ここでWAS120の数のしきい値は、WAS120の数未満の数であり、情報システム10の管理者等により予め設定されてよい。

【0025】

また、しきい値を超える数のWAS120について処理時間が許容範囲内でない場合に（S200:No、S240:Yes）、情報システム10全体の共通資源であるDBS150にボトルネックがあると特定する。ここでボトルネック特定部170は、しきい値を超える数のWAS120について一のアプリケーションプログラム134の処理時間が許容範囲内でない場合（S200:No、S240:Yes）に、処理時間が許容範囲内でないWAS120における、他のアプリケーションプログラム134の処理時間も許容範囲内でないことを更に条件として（S250:Yes）DBS150にボトルネックがあると特定し（S260）、他のアプリケーションプログラム134の処理時間が許容範囲内であれば（S250:No）、当該アプリケーションプログラム134に対応するトランザクションに関連する資源がボトルネックであると特定してもよい（S270）。

【0026】

なお、本実施形態に係るボトルネック特定部170は、以上に示した処理をアプリケーションプログラム134毎又はアプリケーションプログラム134のグループ毎に行う。すなわち、ボトルネック特定部170は、一のアプリケーションプログラム134又は一のグループに属するアプリケーションプログラム134の処理時間が許容範囲内でないことを条件として、当該アプリケーションプログラム134又は当該グループに属するアプリケーションプログラム134に関してWAS120にボトルネックがあると特定する。より具体的には、ボトルネック特定部170は、しきい値以下のWAS120について一のアプリケーションプログラム134又は一のグループに属するアプリケーションプログラム134の処理時間が許容範囲内でない場合に、これらのWAS120にボトルネックがあると特定し、しきい値を超えるWAS120について一のアプリケーションプログラム134又は一のグループに属するアプリケーションプログラム134の処理時間が予め定められた許容範囲内でない場合に、DBS150にボトルネックがあると特定する。

【0027】

以上に示したボトルネック特定部170によれば、処理時間が予め定められた許容範囲内でないWAS120にボトルネックがあると特定することができる。より具体的には、ボトルネック特定部170は、処理時間が許容範囲内でないWAS120の数がしきい値以下であることを条件としてこれらのWAS120にボトルネックがあると特定し、処理時間が許容範囲内でないWAS120の数がしきい値を超える場合にはDBS150にボトルネックがあると特定することができる。すなわち例えばしきい値を1とした場合、ボトルネック特定部170は、処理時間が許容範囲内でないWAS120が1台である場合に当該WAS120にボトルネックがあると特定し、2台以上の場合にDBS150にボトルネックがあると特定することができる。また、例えばしきい値を（WAS120の台数-1）とした場合、全てのWAS120について処理時間が許容範囲内でない場合にDBS150にボトルネックがあると特定し、WAS120の台数未満のWAS120について処理時間が許容範囲内でない場合にこれらのWAS120にボトルネックがあると特定することができる。

【0028】

これにより情報システム10は、ある特定のWAS120のサービスレベルが低下しているのか、それとも情報システム10全体に共通するDBS150がボトルネックとなつて多数のWAS120のサービスレベルが低下しているのかを区別し、適切にボトルネックの存在箇所を特定することができる。ここで、ボトルネック特定部170は、例えばS210又はS250等以外の条件に更に基づいて、より詳細にボトルネックの存在箇所を特定してもよい。

【0029】

以上において、処理時間の許容範囲は、例えば以下に示す予め定められた基準に基づき判断されてよい。すなわち処理時間監視部165は、例えば60秒等の予め定められた期間内にあるアプリケーションプログラム134に対してトランザクションの入力が予め定められた数以上ある場合において、例えば50%等の予め定められた割合以上のトランザクションについて当該アプリケーションプログラム134の処理時間が許容処理時間を超える場合に、当該アプリケーションプログラム134の処理時間が許容範囲を超えると判断する。この場合において、許容処理時間は、例えば過去の1ヶ月間等の期間内におけるピーク時の処理時間の平均値に、標準偏差に基づき定めたオフセット時間を加えて定められる。

【0030】

図3は、本実施形態に係る情報システム10の高負荷判定方法を示す。処理負荷取得部142及び処理負荷取得部156は、監視対象となる複数の資源のそれぞれに対応する「検出条件」により指定された条件が成立した場合に当該資源が高負荷であることを検出する。そして、当該資源の高負荷の検出が「判定条件」により指定された回数等の条件を満たして行われた場合に、当該資源の使用量が許容範囲外となり高負荷な状態であると判定し、当該資源についてのシステム資源高負荷イベントを処理負荷監視部190へ通知する。また、当該資源の高負荷の通知後、当該資源の高負荷の検出が「解除条件」により指定された回数等の条件を満たす間行われなかった場合に、当該資源が高負荷でなくなったことを示すシステム資源高負荷解除イベントを処理負荷監視部190に通知する。これを受けて処理負荷監視部190は、当該データベース資源が高負荷でなくなったことを検出する。

【0031】

処理負荷取得部156は、「ボトルネック箇所の分類」欄が「全体共通資源(DBS)」と示されている複数のデータベースサーバ資源のそれぞれについて、「検出条件」、「判定条件」及び「解除条件」に示した条件を用いて当該資源が高負荷であるか否かを判定する。ここで、本実施形態に係る処理負荷取得部156は、複数のデータベースサーバ資源として、プロセッサ、ディスク、メモリ、データベースバッファ等を含む複数のデータベースサーバ資源のそれぞれについて、条件を用いて高負荷であるか否かを判定する。

【0032】

プロセッサ、データベースを格納するディスク、DB152のアクセスログを格納するディスク、データベースとアクセスログ以外のファイルを格納するディスクに関して、処理負荷取得部156は、当該データベース資源の使用率が「n秒間の間平均x%以上」である場合に高負荷であることを検出する。そして処理負荷取得部156は、この高負荷の状態がモニタリングするタイミングにおいてm回連続して検出された場合に、当該データベース資源が高負荷であると判定し、システム資源高負荷イベントを処理負荷監視部190に通知する。また処理負荷取得部156は、システム資源高負荷イベントを処理負荷監視部190に通知した後、高負荷の状態をf回連続して検出しなかった場合にシステム資源高負荷解除イベントを処理負荷監視部190に通知する。

【0033】

また、処理負荷取得部156は、DBS150が搭載する実メモリに関して、ページングが「n秒間の間平均x回以上」発生している場合に、当該メモリの使用量が多くスワップが頻発しているとして高負荷であることを検出する。また、DBエンジン154が使用

するデータベースバッファに関して、当該バッファのミス率が「 n 秒間の間平均 $x\%$ 以上」の場合に、当該バッファが不足しているとして高負荷であることを検出する。処理負荷取得部156は、これらのデータベースサーバ資源についてのシステム資源高負荷イベント及びシステム資源高負荷解除イベントの通知を、プロセッサ等のシステム資源高負荷イベント及びシステム資源高負荷解除イベントの通知と同様に行う。

【0034】

複数のWAS120のそれぞれの処理負荷取得部142は、「ボトルネック箇所の分類」欄が「WAS」と示されている複数のアプリケーションサーバ資源のそれぞれについて、「検出条件」、「判定条件」及び「解除条件」に示した条件を用いて当該資源が高負荷であるか否かを判定する。ここで、本実施形態に係る処理負荷取得部142は、複数のデータベースサーバ資源として、プロセッサ、ディスク、メモリ、JVM (Java Virtual Machine) メモリ、及びDBコネクション等を含む複数のアプリケーションサーバ資源のそれぞれについて、「検出条件」、「判定条件」及び「解除条件」に示した条件を用いて高負荷であるか否かを判定する。

【0035】

プロセッサ、ディスク、及びメモリに関して、処理負荷取得部142は、処理負荷取得部156と同様にして高負荷の検出及びイベントの通知を行う。また処理負荷取得部142は、WEBアプリケーションサーバ部130を構成する各種のプログラム又はモジュールを実行するJVMが「ガベージコレクションを n 秒間の間に x 回以上」実行した場合に、JVMに割り当てたメモリが不足しているとして高負荷であることを検出する。そして処理負荷取得部142は、JVMメモリについてのシステム資源高負荷イベント及びシステム資源高負荷解除イベントの通知を、プロセッサ等のシステム資源高負荷イベント及びシステム資源高負荷解除イベントの通知と同様に行う。

【0036】

また、処理負荷取得部142は、DBコネクションに関して、DBコネクション139を使用する多重度が上限に達した場合に、DBコネクションが不足しているとして高負荷であることを検出する。処理負荷取得部142は、一旦DBコネクションが高負荷であることを検出すると、システム資源高負荷イベントを処理負荷監視部190に通知し、当該WAS120を停止するまで高負荷でなくなったことを処理負荷監視部190に通知しない。

【0037】

また、処理負荷取得部142は、「ボトルネック箇所の分類」欄が「トランザクションに関連する資源」と示されているWAS120のプロセッサ、DBS150のプロセッサ、及びDBS150のディスク等の各資源について、「検出条件」、「判定条件」及び「解除条件」に示した条件を用いて当該資源がアプリケーションプログラム134の処理内容の影響により高負荷となっているか否かを判定する。

【0038】

処理負荷取得部142は、WAS120のプロセッサに関して、トランザクションに対応してアプリケーションプログラム134を実行する処理時間全体のうち、DBS150によるDB152のアクセスに要する処理時間の比率が $x\%$ 以下である場合に、WAS120のプロセッサによる処理時間の割合が高いとして高負荷を検出する。そして処理負荷取得部142は、高負荷を検出するとシステム資源高負荷イベントを処理負荷監視部190に通知し、検出条件に示された状態が n 秒間発生しない場合にシステム資源高負荷解除イベントを処理負荷監視部190に通知する。

【0039】

また、処理負荷取得部142は、DBS150のプロセッサに関して、トランザクションに対応してアプリケーションプログラム134を実行する処理時間全体のうちDBS150によるDB152のアクセスに要する処理時間の比率が $x\%$ 以上である場合に、DBS150のプロセッサによる処理時間の割合が高いとして高負荷を検出する。そして処理負荷取得部142は、高負荷を検出するとシステム資源高負荷イベントを処理負荷監視部

190に通知し、高負荷の検出がn秒間発生しない場合にシステム資源高負荷解除イベントを処理負荷監視部190に通知する。これを受けて処理負荷監視部190は、当該システム資源高負荷イベントの通知を受け、かつ、n秒以内にDBS150のプロセッサがボトルネックであるとDBS150により判定されていた場合に、トランザクションに対応してDBS150のプロセッサが高負荷であると判定する。また、処理負荷取得部142は、DBS150のディスクについても、DBS150のプロセッサと同様にして、トランザクションに対応してDBS150のディスクが高負荷であると判定する。

【0040】

処理負荷監視部190は、複数のWAS120及びDBS150のそれぞれについて、当該サーバが有する複数の資源のそれぞれが高負荷であるか否かを、以上に例示したシステム資源高負荷イベント及びシステム資源高負荷解除イベントに基づいて検出する。ここで処理負荷監視部190は、複数の資源のそれぞれに対応する「検出条件」、「判定条件」及び「解除条件」を、例えば負荷制御サーバ160が有する記憶装置に格納したファイルとして保持し、情報システム10の管理者等の指示に基づき変更可能としてもよい。

【0041】

また、ボトルネック特定部では、処理負荷取得部142により、アプリケーションプログラム134単位にシステム資源高負荷イベントが通知されている場合、および、処理負荷取得部156により、DB処理要求I/F部153単位にシステム資源高負荷イベントが通知されている場合には、より詳細にボトルネックの原因となるアプリケーションプログラム134を特定してもよい。

【0042】

図4は、本実施形態に係る負荷制御情報格納部177に格納される負荷制御情報の一例を示す。負荷制御情報格納部177は、WAS120における複数のアプリケーションサーバ資源、及び、DBS150における複数のデータベースサーバ資源のそれぞれに対応して、当該資源を有するサーバにボトルネックがあると特定され、かつ、当該資源の使用量が予め定められた許容範囲内でない場合に負荷制御部175が行うべき負荷制御処理を指定する負荷制御情報を格納する。ここで負荷制御情報格納部177は、少なくとも1つの負荷制御情報として、複数の負荷制御処理と複数の負荷制御処理の優先順位とを指定する情報を格納してもよい。

【0043】

負荷制御情報取得部179は、ボトルネックがあると特定されたサーバが有する資源の使用量が予め定められた許容範囲内でない場合に、当該資源がボトルネックであると特定する。そして、負荷制御情報取得部179は、当該資源がボトルネックであると特定された場合に、当該資源に対応して負荷制御情報格納部177に格納された負荷制御情報を取得する。より具体的には、負荷制御情報取得部179は、WAS120にボトルネックがあると特定された場合に、当該WAS120が有する複数のアプリケーションサーバ資源のうち、使用量が許容範囲内でないと処理負荷監視部190により判定されたアプリケーションサーバ資源に対応して、負荷制御情報格納部177に格納された負荷制御情報を取得する。また、負荷制御情報取得部179は、DBS150にボトルネックがあると特定された場合に、複数のデータベースサーバ資源のうち、使用量が許容範囲内でないと処理負荷監視部190により判定されたデータベースサーバ資源に対応して負荷制御情報格納部177に格納された負荷制御情報を取得する。

【0044】

そして、負荷制御処理部181は、ボトルネック特定部170によりボトルネックがあると特定されたサーバのボトルネックを解消するべく、負荷制御情報取得部179により取得された負荷制御情報により指定される負荷制御処理を行う。ここで負荷制御処理部181は、ある負荷制御情報により指定された一の負荷制御処理を実行した後、ボトルネックがあると特定されたサーバのボトルネックが予め設定された設定期間の間に解消しないことを検出した場合に、当該負荷制御情報より指定される一の負荷制御処理より優先順位が低い他の負荷制御処理を実行する。これにより、負荷制御処理部181は、1種類の負

荷制御処理のみではボトルネックが解消されない場合においても、他の負荷制御処理を更に行うことによってより効果的にボトルネックを解消することができる。

【0045】

本実施形態に係る負荷制御処理部181は、以下に示す負荷制御処理を行う。

(1) 負荷制御処理1

負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134より優先度が低いグループのアプリケーションプログラム134を実行する多重度を低下させる。より具体的には、負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134より低い優先度の複数のグループ識別子を、ボトルネック特定部170から負荷制御情報取得部179を介して取得する。そして、負荷制御処理部181は、取得した複数のグループ識別子を指定して、当該アプリケーションプログラム134より優先度が低い他のアプリケーションプログラム134の多重度を低下させるようにサーバ内負荷制御部144に指示する。この指示を受けてサーバ内負荷制御部144は、多重度制御部132が他のアプリケーションプログラム134を並行して動作させる数の最大値を低下させることにより、当該他のアプリケーションプログラム134を実行する多重度を低下させる。

DBS150にボトルネックがあると特定された場合において負荷制御処理1を実行する場合、負荷制御処理部181は、情報システム10が有する複数のWAS120のそれぞれについて上記の負荷制御処理1を実行する。これにより負荷制御処理部181は、全てのWAS120がトランザクションを受け入れる多重度を低減させて情報システム10におけるトランザクションの総処理量を低減させることができ、DB152のアクセス量を低減させることができる。

【0046】

(2) 負荷制御処理2

負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134より優先度が低いグループのアプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を更に低下させる。より具体的には、負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134より低い優先度の複数のグループ識別子を、ボトルネック特定部170から負荷制御情報取得部179を介して取得する。そして、負荷制御処理部181は、取得した複数のグループ識別子を指定して、当該アプリケーションプログラム134より優先度が低い他のアプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を低下させるようにサーバ内負荷制御部158に指示する。この指示を受けてサーバ内負荷制御部158は、対応するDB処理要求I/F部153のプロセス実行優先度を低下させる。

【0047】

(3) 負荷制御処理3

ボトルネック特定部により、ボトルネックが存在するサーバにおけるボトルネックとなっている資源の処理負荷を高くしているアプリケーションプログラム134がシステム資源高負荷イベントにより特定できている場合、負荷制御処理部181は、当該アプリケーションプログラム134のグループを実行する多重度を低下させる。より具体的には、負荷制御処理部181は、ボトルネックの原因となるアプリケーションプログラム134のグループ識別子を、ボトルネック特定部170から負荷制御情報取得部179を介して取得する。そして、負荷制御処理部181は、取得したグループ識別子を指定して、当該アプリケーションプログラム134の多重度を低下させるようにサーバ内負荷制御部144に指示する。この指示を受けてサーバ内負荷制御部144は、多重度制御部132がアプリケーションプログラム134を並行して動作させる数の最大値を低下させることにより、当該他のアプリケーションプログラム134を実行する多重度を低下させる。

DBS150にボトルネックがあると特定された場合において負荷制御処理3を実行する場合、負荷制御処理部181は、情報システム10が有する複数のWAS120のそれ

それぞれについて上記の負荷制御処理3を実行する。これにより負荷制御処理部181は、全てのWAS120がトランザクションを受け入れる多重度を低減させて情報システム10におけるトランザクションの総処理量を低減させることができ、DB152のアクセス量を低減させることができる。

【0048】

(4) 負荷制御処理4

負荷制御処理部181は、ボトルネック特定部170により、ボトルネックが存在するサーバにおけるボトルネックとなっている資源の処理負荷を高くしているアプリケーションプログラム134がシステム資源高負荷イベントにより特定できている場合、当該アプリケーションプログラム134を実行するグループのアプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を低下させる。より具体的には、負荷制御処理部181は、ボトルネックの原因となるアプリケーションプログラム134のグループ識別子を、ボトルネック特定部170から負荷制御情報取得部179を介して取得する。そして、負荷制御処理部181は、取得したグループ識別子を指定して、当該アプリケーションプログラム134のグループのアプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を低下させるようにサーバ内負荷制御部158に指示する。この指示を受けてサーバ内負荷制御部158は、対応するDB処理要求I/F部のプロセス実行優先度を低下させる。

【0049】

(5) 負荷制御処理5

負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134を実行する多重度を低下させる。より具体的には、負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134のグループ識別子を、ボトルネック特定部170から負荷制御情報取得部179を介して取得する。そして、負荷制御処理部181は、当該アプリケーションプログラム134のグループ識別子を指定して、当該アプリケーションプログラム134の多重度を低下させるようにサーバ内負荷制御部144に指示する。この指示を受けてサーバ内負荷制御部144は、多重度制御部132が当該アプリケーションプログラム134を並行して動作させる数の最大値を低下させることにより、当該アプリケーションプログラム134を実行する多重度を低下させる。

DBS150にボトルネックがあると特定された場合において負荷制御処理3を実行する場合、負荷制御処理部181は、負荷制御処理1の場合と同様にして、情報システム10が有する複数のWAS120のそれぞれについて上記の負荷制御処理3を実行する。

【0050】

(6) 負荷制御処理6

負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を低下させる。より具体的には、負荷制御処理部181は、処理時間が許容範囲内でないと判断されたアプリケーションプログラム134のグループ識別子を、ボトルネック特定部170から負荷制御情報取得部179を介して取得する。そして、負荷制御処理部181は、当該アプリケーションプログラム134のグループ識別子を指定して、当該アプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を低下させるようにサーバ内負荷制御部158に指示する。この指示を受けてサーバ内負荷制御部158は、対応するDB処理要求I/F部153のプロセス実行優先度を低下させる。

【0051】

本実施形態に係る負荷制御情報格納部177は、DBS150のボトルネックに関して、複数のデータベースサーバ資源のいずれの使用量が許容範囲外となったかに応じて、図4に示した負荷制御情報を格納する。

より具体的には、プロセッサの使用量が許容範囲外となった場合に対応して、負荷制御情報格納部177は、負荷制御処理3、1、4、2、5及び6をこの優先順位で行うべき

ことを示す負荷制御情報を格納する。また、データベースを格納しているディスクの使用量が許容範囲外となった場合に対応して、負荷制御情報格納部177は、負荷制御処理2、1、6、及び5をこの優先順位で行うべきことを示す負荷制御情報を格納する。また、アクセスログを格納するディスクを格納するディスクの使用量が許容範囲外となった場合に対応して、負荷制御情報格納部177は、負荷制御処理1、2、5、及び6をこの優先順位で行うべきことを示す負荷制御情報を格納する。また、データベースおよびアクセスログ以外を格納するディスク、メモリ、又はデータベースバッファの使用量が許容範囲外となった場合に対応して、負荷制御情報格納部177は、負荷制御処理1及び5をこの優先順位で行うべきことを示す負荷制御情報を格納する。

【0052】

これらの負荷制御情報を用いて、負荷制御部175は、例えば、DBS150にボトルネックがあると特定され、さらにプロセッサの使用量が許容範囲の使用量を超えておりプロセッサがボトルネックであると特定された場合において、DBS150のプロセッサ資源の使用量が許容範囲内でないアプリケーションプログラム134が特定できている場合は、負荷制御処理3を実行し、当該アプリケーションプログラム134及び当該アプリケーションプログラム134と同一のグループに属するアプリケーションプログラム134をWAS120により実行する多重度を低下させる。そして、この負荷制御処理によってもDBS150のボトルネックが解消しない場合に、負荷制御部175は、前記負荷制御処理1を実行することにより、更に当該アプリケーションプログラム134より優先度が低いグループに属する他のアプリケーションプログラム134をWAS120により実行する多重度を低下させることができる。これにより負荷制御部175は、まず、プロセッサを大量に消費しているアプリケーションプログラム134の処理量を減らしてDBS150の負荷低減を試み、これによって解決しない場合に当該アプリケーションプログラム134より優先度が低いアプリケーションプログラム134の処理量を減らしていくことにより、段階的にDBS150の負荷を減らすことができる。

【0053】

そしてこれらの負荷制御処理によってもボトルネックが解消しない場合には、負荷制御部175は、負荷制御処理4、2を更に順に実行することにより、DBS150上で処理している要求への他の処理からの影響を軽減し、サービスレベルの改善を試みる。当初処理時間が許容範囲内でないと判定されたアプリケーションプログラム134自身の問題、又は、より優先度の高いアプリケーションプログラム134の問題でなければ、ここまでの負荷制御でボトルネックが解消され则认为られる。しかし、それでもレスポンスが改善しない場合には、負荷制御処理5により、サービスレベルが低下したアプリケーションプログラム134自体をWAS120により実行する多重度を減らして、DBS150の負荷をさらに減らすことができる。また、負荷制御処理5によってもレスポンスが改善しない場合には、負荷制御部175は、負荷制御処理6を実行することにより、アプリケーションプログラム134の要求に基づきDB152をアクセスする処理の優先度を低下させ、DBS150上で処理している要求が他の処理へ与える影響を軽減できる。

【0054】

一方、負荷制御部175は、例えばDBS150にボトルネックがあると特定され、さらにデータベースを格納するディスクがボトルネックであると特定された場合において、DBS150の当該ディスク資源の使用量が許容範囲内でないアプリケーションプログラム134が特定できている場合は、負荷制御処理2を実行し、当該アプリケーションプログラム134より優先度が低いグループに属するアプリケーションプログラム134からの要求に基づきDB152をアクセスする処理の優先度を低下させる。ここでデータベースを格納するディスクが複数存在する場合、負荷制御部175は、負荷制御処理2において、当該アプリケーションプログラム134より優先度が低いグループに属し、当該アプリケーションプログラム134が使用するデータベースが格納されたディスクを使用するアプリケーションプログラム134からの要求に基づきDB152をアクセスする処理の優先度を低下させることが望ましい。そして、負荷制御処理2によってもDBS150の

ボトルネックが解消しない場合に、負荷制御部 175 は、負荷制御処理 1、6、及び 5 を DBS 150 のボトルネックが解消されるまで順次実行する。

【0055】

また、本実施形態に係る負荷制御情報格納部 177 は、WAS 120 のボトルネックに関して、複数のアプリケーションサーバ資源のいずれの使用量が許容範囲外となったかに応じて、図 4 に示した負荷制御情報を格納する。

より具体的には、プロセッサが許容範囲外となった場合に対応して、負荷制御情報格納部 177 は、負荷制御処理 3、1、及び 5 をこの優先順位で行うべきことを示す負荷制御情報を格納する。また、メモリ、ディスク、JVM メモリ、及び DB コネクションの使用量の使用量が許容範囲外となった場合に対応して、負荷制御情報格納部 177 は、負荷制御処理 1、及び 5 を行うべきことを示す負荷制御情報を格納する。

【0056】

これらの負荷制御情報を用いて、負荷制御部 175 は、例えば WAS 120 にボトルネックがあると特定され、さらにプロセッサの使用量が許容範囲の使用量を超えておりプロセッサがボトルネックであると特定された場合において、WAS 120 のプロセッサ資源の使用量が許容範囲内でないアプリケーションプログラム 134 が特定できている場合は、負荷制御処理 3 を実行し、当該のアプリケーションプログラム 134 及び当該アプリケーションプログラム 134 と同一のグループに属するアプリケーションプログラム 134 を WAS 120 により実行する多重度を低下させる。そして、この負荷制御処理によっても WAS 120 のボトルネックが解消しない場合に、負荷制御部 175 は、負荷制御処理 1 を実行し、更に当該アプリケーションプログラム 134 より優先度が低いグループに属する他のアプリケーションプログラム 134 を WAS 120 により実行する多重度を低下させることができる。

【0057】

当初処理時間が許容範囲内でないと判定されたアプリケーションプログラム 134 自身の問題、又は、より優先度の高いアプリケーションプログラム 134 の問題でなければ、ここまでの負荷制御でボトルネックが解消されると考えられる。しかし、それでもレスポンスが改善しない場合には、負荷制御部 175 は、負荷制御処理 5 を実行することにより、サービスレベルが低下したアプリケーションプログラム 134 自体を WAS 120 が実行する多重度を減らして、DBS 150 の負荷をさらに減らすことができる。

【0058】

また、本実施形態に係る負荷制御情報格納部 177 は、トランザクションの内容やアプリケーションプログラム 134 の処理内容に関連する資源のボトルネックに関して、WAS 120 のプロセッサの使用量が許容範囲外になった場合に、負荷制御処理 5 を行うべきことを示す負荷制御情報を格納する。また、負荷制御情報格納部 177 は、トランザクションの内容やアプリケーションプログラム 134 の処理内容に関連する資源のボトルネックに関して、DBS 150 のプロセッサ又はディスクの使用量が許容範囲外になった場合に、負荷制御処理 5 及び 6 をこの優先順位で行うべきことを示す負荷制御情報を格納する。

【0059】

図 5 は、本発明の実施形態に係るコンピュータ 900 のハードウェア構成の一例を示す。本実施形態に係るコンピュータ 900 は、負荷制御プログラムを実行することにより負荷制御サーバ 160 として機能する。コンピュータ 900 は、ホスト・コントローラ 1082 により相互に接続される CPU 1000、RAM 1020、グラフィック・コントローラ 1075、及び表示装置 1080 を有する CPU 周辺部と、入出力コントローラ 1084 によりホスト・コントローラ 1082 に接続される通信インターフェイス 1030、記憶装置 1040、及び CD-ROM ドライブ 1060 を有する入出力部と、入出力コントローラ 1084 に接続される ROM 1010、フレキシブルディスク・ドライブ 1050、及び入出力チップ 1070 を有するレガシー入出力部とを備える。

【0060】

ホスト・コントローラ1082は、RAM1020と、高い転送レートでRAM1020をアクセスするCPU1000及びグラフィック・コントローラ1075とを接続する。CPU1000は、ROM1010及びRAM1020に格納されたプログラムに基づいて動作し、各部の制御を行う。グラフィック・コントローラ1075は、CPU1000等がRAM1020内に設けたフレーム・バッファ上に生成する画像データを取得し、表示装置1080上に表示させる。これに代えて、グラフィック・コントローラ1075は、CPU1000等が生成する画像データを格納するフレーム・バッファを、内部に含んでもよい。

【0061】

入出力コントローラ1084は、ホスト・コントローラ1082と、比較的高速な入出力装置である通信インターフェイス1030、記憶装置1040、CD-ROMドライブ1060を接続する。通信インターフェイス1030は、ネットワークを介して他の装置と通信する。記憶装置1040は、コンピュータ900内のCPU1000が使用するプログラム及びデータを格納する。CD-ROMドライブ1060は、CD-ROM1095からプログラム又はデータを読み取り、RAM1020を介して記憶装置1040に提供する。

【0062】

また、入出力コントローラ1084には、ROM1010と、フレキシブルディスク・ドライブ1050や入出力チップ1070等の比較的低速な入出力装置とが接続される。ROM1010は、コンピュータ900が起動時に実行するブート・プログラムや、コンピュータ900のハードウェアに依存するプログラム等を格納する。フレキシブルディスク・ドライブ1050は、フレキシブルディスク1090からプログラム又はデータを読み取り、RAM1020を介して記憶装置1040に提供する。入出力チップ1070は、フレキシブルディスク・ドライブ1050や、例えばパラレル・ポート、シリアル・ポート、キーボード・ポート、マウス・ポート等を介して各種の入出力装置を接続する。

【0063】

RAM1020を介して記憶装置1040に提供されるプログラムは、フレキシブルディスク1090、CD-ROM1095、又はICカード等の記録媒体に格納されて利用者によって提供される。WAS120をHTTPサーバ部125、WEBアプリケーションサーバ部130、処理時間取得部140、処理負荷取得部142、及びサーバ内負荷制御部144として機能させるプログラムは、記録媒体から読み出され、通信インターフェイス1030を介してWAS120にインストールされ、WAS120として用いられるサーバコンピュータにおいて実行される。また、DBS150をDB152、DB処理要求I/F部153、処理負荷取得部156及びサーバ内負荷制御部158として機能させるプログラムは、記録媒体から読み出され、通信インターフェイス1030及び負荷制御サーバ160を介してDBS150にインストールされ、DBS150として用いられるサーバコンピュータにおいて実行される。また、コンピュータ900を負荷制御サーバ160として機能させる負荷制御プログラムは、記録媒体から読み出され、RAM1020を介してコンピュータ900内の記憶装置1040にインストールされ、CPU1000において実行される。

【0064】

WAS120にインストールされるプログラムは、HTTPサーバプログラムと、多重度制御モジュール及びDBアクセス制御モジュールを有するWEBアプリケーションサーバプログラムと、処理時間取得モジュールと、処理負荷取得モジュールと、サーバ内負荷制御モジュールとを備える。これらのプログラム及びモジュールは、WAS120を、HTTPサーバ部125と、多重度制御部132及びDBアクセス制御部136を有するWEBアプリケーションサーバ部130と、処理時間取得部140と、処理負荷取得部142と、サーバ内負荷制御部144としてそれぞれ機能させる。

【0065】

DBS150にインストールされるプログラムは、DBエンジンプログラムと、処理負

荷取得モジュールと、サーバ内負荷制御モジュールとを備える。これらのプログラム及びモジュールは、DBS150を、DBエンジン154と、処理負荷取得部156と、サーバ内負荷制御部158としてそれぞれ機能させる。

【0066】

コンピュータ900にインストールされ、コンピュータ900を負荷制御サーバ160として機能させる負荷制御プログラムは、処理時間監視モジュールと、ボトルネック特定モジュールと、処理負荷監視モジュールと、負荷制御情報取得モジュール及び負荷制御処理モジュールを有する負荷制御モジュールとを備える。これらのプログラム及びモジュールは、コンピュータ900を、処理時間監視部165と、ボトルネック特定部170と、処理負荷監視部190と、負荷制御情報取得部179及び負荷制御処理部181を有する負荷制御部175としてそれぞれ機能させる。

【0067】

以上に示したプログラム又はモジュールは、外部の記憶媒体に格納されてもよい。記憶媒体としては、フレキシブルディスク1090、CD-ROM1095の他に、DVDやPD等の光学記録媒体、MD等の光磁気記録媒体、テープ媒体、ICカード等の半導体メモリ等を用いることができる。また、専用通信ネットワークやインターネットに接続されたサーバシステムに設けたハードディスク又はRAM等の記憶装置を記録媒体として使用し、ネットワークを介してプログラムをコンピュータ900に提供し、コンピュータ900から負荷制御サーバ160を介してWAS120又はWEBアプリケーションサーバ部130へ提供してもよい。

【0068】

以上に示した情報システム10によれば、負荷制御情報格納部177は、ボトルネックの存在箇所及び高負荷と判定された資源に対応して、適切な負荷制御処理を格納することができる。そして負荷制御部175は、負荷制御情報格納部177に格納された負荷制御情報に基づいて、WAS120上で実行されるアプリケーションプログラム134の多重度を変更したり、DB152をアクセスする処理の優先度を変更したりする等の負荷制御処理を適切に行うことができる。

【0069】

更に情報システム10によれば、複数のアプリケーションプログラム134のそれぞれの処理時間に応じて、複数のアプリケーションプログラム134のそれぞれが実行される多重度及びDBS150におけるアクセスの優先順位を適切に変更することができる。これにより、多種多様な業務システム等を併せて情報システム10で稼働させる場合においても、あるアプリケーションプログラム134が異常にシステム資源を消費した場合の影響で、他のアプリケーションプログラム134のサービスレベルが低下しても、自動的に検知し、短時間で自動的に復旧させることができ、それぞれのアプリケーションプログラム134を別個の情報システム10で実行した場合と同等の頑強さを有する情報システム10を実現することができる。

【0070】

以上、本発明を実施の形態を用いて説明したが、本発明の技術的範囲は上記実施の形態に記載の範囲には限定されない。上記実施の形態に、多様な変更または改良を加えることが可能であることが当業者に明らかである。その様な変更または改良を加えた形態も本発明の技術的範囲に含まれ得ることが、特許請求の範囲の記載から明らかである。

【0071】

以上に説明した実施形態によれば、以下の各項目に示す情報システム、負荷制御方法、負荷制御プログラム及び記録媒体が実現される。

【0072】

(項目1) 端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システムであって、前記複数のアプリケーションサーバのそれぞれについて、当該アプリ

ケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間を監視する処理時間監視部と、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあると特定するボトルネック特定部と、ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる負荷制御部とを備える情報システム。

【0073】

(項目2) 前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記アプリケーションサーバ数未満のしきい値以下の数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でないことを条件として、当該アプリケーションサーバにボトルネックがあると特定し、前記しきい値を超える数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でない場合に、前記データベースサーバにボトルネックがあると特定し、前記負荷制御部は、当該アプリケーションサーバにボトルネックがあると特定された場合に、当該アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させ、前記データベースサーバにボトルネックがあると特定された場合に、前記複数のアプリケーションサーバのそれぞれについて前記アプリケーションプログラムを実行する多重度を低下させる項目1記載の情報システム。

【0074】

(項目3) 前記複数のアプリケーションサーバのそれぞれは、当該アプリケーションサーバ上で実行されるべき優先度が設定された複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でないことを条件として、当該アプリケーションサーバにボトルネックがあると特定し、前記負荷制御部は、ボトルネックがあると特定された前記アプリケーションサーバ上で前記一の前記アプリケーションプログラムより優先度が低い前記アプリケーションプログラムを実行する多重度を低下させる項目1記載の情報システム。

【0075】

(項目4) 前記複数のアプリケーションサーバのそれぞれは、当該アプリケーションサーバ上で実行されるべき優先度が設定された複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、予め定められたしきい値を超える数の前記アプリケーションサーバについて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でない場合に、前記データベースサーバにボトルネックがあると特定し、前記負荷制御部は、ボトルネックがあると特定された前記データベースサーバにおける、前記一の前記アプリケーションプログラムより優先度が低い前記アプリケーションプログラムの要求に基づき前記データベースをアクセスする処理の優先度を低下させる項目1記載の情報システム。

【0076】

(項目5) 前記複数のアプリケーションサーバのそれぞれは、当該アプリケーションサーバ上で実行されるべき優先度が設定された複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーション

ンプログラムを実行し、前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、予め定められたしきい値を超える数の前記アプリケーションサーバについて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でない場合に、前記データベースサーバにボトルネックがあると特定し、前記負荷制御部は、ボトルネックがあると特定された前記データベースサーバにおける、前記一のアプリケーションプログラムからの要求に基づき前記データベースをアクセスする処理の優先度を低下させる項目 1 記載の情報システム。

【0077】

(項目 6) 前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが有する、前記アプリケーションプログラムにより使用される資源であるアプリケーションサーバ資源の使用量を監視する処理負荷監視部を更に備え、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記アプリケーションサーバ数未満のしきい値以下の数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でない場合に、当該アプリケーションサーバにボトルネックがあると特定し、前記負荷制御部は、ボトルネックがあると特定された前記アプリケーションサーバにおける前記アプリケーションサーバ資源の使用量が予め定められた許容範囲内でないことを条件として、ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる項目 1 記載の情報システム。

【0078】

(項目 7) 前記処理負荷監視部は、前記データベースサーバが有する、前記アプリケーションプログラムが前記データベースをアクセスする場合に使用される資源であるデータベースサーバ資源の使用量を更に監視し、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記しきい値を超える数の前記アプリケーションサーバについて前記処理時間が予め定められた許容範囲内でないことを条件として、前記データベースサーバにボトルネックがあると特定し、前記負荷制御部は、ボトルネックがあると特定された前記データベースサーバにおける前記データベースサーバ資源の使用量が予め定められた許容範囲内でない場合に、前記複数のアプリケーションサーバのそれぞれについて当該アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる項目 6 記載の情報システム。

【0079】

(項目 8) 前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが有する、前記アプリケーションプログラムにより使用される資源である複数のアプリケーションサーバ資源の使用量と、前記データベースサーバが有する、前記アプリケーションプログラムが前記データベースをアクセスする場合に使用される資源である複数のデータベースサーバ資源の使用量とを監視する処理負荷監視部を更に備え、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記複数のアプリケーションサーバ及び前記データベースサーバのうち少なくとも 1 つの前記アプリケーションサーバ又は前記データベースサーバにボトルネックがあると特定し、前記負荷制御部は、前記アプリケーションサーバにおける前記複数のアプリケーションサーバ資源、及び、前記データベースサーバにおける前記複数のデータベースサーバ資源のそれぞれに対応して、当該資源を有するサーバにボトルネックがあると特定され、かつ、当該資源の使用量が予め定められた許容範囲内でない場合に前記負荷制御部が行うべき負荷制御処理を指定する負荷制御情報を格納する負荷制御情報格納部と、ボトルネックがあると特定されたサーバが有する資源の使用量が予め定められた許容範囲内でない場合に、当該資源がボトルネックであると特定し、前記アプリケーションサーバ資源がボトルネックであると特定された場合に、当該資源に対応して前記負荷

制御情報格納部に格納された前記負荷制御情報を取得し、前記データベースサーバ資源がボトルネックであると特定された場合に、当該資源に対応して前記負荷制御情報格納部に格納された前記負荷制御情報を取得する負荷制御情報取得部と、ボトルネックがあると特定されたサーバのボトルネックを解消するべく、前記負荷制御情報により指定される前記負荷制御処理を実行する負荷制御処理部とを有する項目 1 記載の情報システム。

【0080】

(項目 9) 前記複数のアプリケーションサーバのそれぞれは、複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、前記処理負荷監視部は、前記データベースサーバについて、当該データベースサーバが有する前記データベースサーバ資源であるプロセッサ及びデータベースの格納されているディスクの使用量を監視し、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、前記しきい値を超える数の前記アプリケーションサーバについて、一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でないことを条件として、前記データベースサーバにボトルネックがあると特定し、前記負荷制御情報格納部は、前記データベースサーバにおける前記アプリケーション資源であるプロセッサに対応する第 1 の前記負荷制御情報と、データベースの格納されているディスクに対応する第 2 の前記負荷制御情報とを格納し、前記負荷制御情報取得部は、ボトルネックがあると特定された前記データベースサーバが有するプロセッサの使用量が予め定められた使用量を超える場合に前記第 1 の負荷制御情報を取得し、ボトルネックがあると特定された前記データベースサーバが有するデータベースの格納されているディスクの使用量が予め定められた使用量を超える場合に前記第 2 の負荷制御情報を取得し、前記負荷制御処理部は、前記負荷制御情報取得部から前記第 1 の負荷制御情報が取得された場合に、前記複数のアプリケーションサーバのそれぞれについて前記一のアプリケーションプログラムを実行する多重度を低下させ、前記負荷制御情報取得部から前記第 2 の負荷制御情報が取得された場合に、前記データベースサーバにおいて当該制御情報で指定された当該ディスクを使用するアプリケーションプログラムのグループの中で、前記一のアプリケーションプログラムより優先度が低い前記アプリケーションプログラムからの要求に基づき前記データベースをアクセスする処理の優先度を低下させる項目 8 記載の情報システム。

【0081】

(項目 10) 前記負荷制御情報格納部に格納された一の前記負荷制御情報は、複数の前記負荷制御処理と前記複数の負荷制御処理の優先順位とを指定する情報であり、前記負荷制御処理部は、前記一の負荷制御情報により指定された一の前記負荷制御処理を実行した後、ボトルネックがあると特定されたサーバのボトルネックが予め設定された設定期間の間に解消しないことを検出した場合に、前記一の負荷制御情報より指定される前記一の負荷制御処理より優先順位が低い他の前記負荷制御処理を実行する項目 8 記載の情報システム。

【0082】

(項目 11) 前記複数のアプリケーションサーバのそれぞれは、複数種類の前記アプリケーションプログラムのうち、前記端末から受信した前記トランザクションの種類に対応する前記アプリケーションプログラムを実行し、前記処理時間監視部は、前記複数のアプリケーションサーバのそれぞれについて、前記アプリケーションプログラム毎の前記トランザクションの処理時間を監視し、前記ボトルネック特定部は、前記複数のアプリケーションサーバのそれぞれについての前記アプリケーションプログラム毎の前記処理時間の監視結果に基づいて、前記アプリケーションサーバ数未満のしきい値以下の前記アプリケーションサーバについて一の前記アプリケーションプログラムの前記処理時間が予め定められた許容範囲内でない場合に、当該アプリケーションサーバにボトルネックがあると特定し

、前記負荷制御部は、ボトルネックがあると特定された前記アプリケーションサーバ上で前記一のアプリケーションプログラムより低い優先度に設定された前記アプリケーションプログラムを実行する多重度を低下させ、当該アプリケーションサーバのボトルネックが前記設定期間の間に解消しないことを検出した場合に、当該アプリケーションサーバ上で前記一のアプリケーションプログラムを実行する多重度を更に低下させる項目 8 記載の情報システム。

【0083】

(項目 12) 端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システムをコンピュータにより制御する負荷制御方法であって、前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間をコンピュータにより監視する処理時間監視段階と、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあるとコンピュータにより特定するボトルネック特定段階と、ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度をコンピュータにより低下させる負荷制御段階とを備える負荷制御方法。

【0084】

(項目 13) 端末から受信したトランザクションに応じてアプリケーションプログラムをそれぞれ実行する複数のアプリケーションサーバと、前記複数のアプリケーションサーバのそれぞれの要求に基づきデータベースをアクセスするデータベースサーバとを備える情報システムをコンピュータにより制御させる負荷制御プログラムであって、前記コンピュータを、前記複数のアプリケーションサーバのそれぞれについて、当該アプリケーションサーバが受信した前記トランザクションを前記アプリケーションプログラムが処理する処理時間を監視する処理時間監視部と、前記複数のアプリケーションサーバのそれぞれについての前記処理時間の監視結果に基づいて、前記処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあると特定するボトルネック特定部と、ボトルネックがあると特定された前記アプリケーションサーバ上で前記アプリケーションプログラムを実行する多重度を低下させる負荷制御部として機能させる負荷制御プログラム。

【0085】

(項目 14) 項目 13 に記載の負荷制御プログラムを記録した記録媒体。

【図面の簡単な説明】

【0086】

【図 1】 本発明の実施形態に係る情報システム 10 の構成を示す。

【図 2】 本発明の実施形態に係る情報システム 10 のボトルネック特定フローを示す。

。 【図 3】 本発明の実施形態に係る情報システム 10 の高負荷判定方法を示す。

【図 4】 本発明の実施形態に係る負荷制御情報格納部 177 に格納される負荷制御情報の一例を示す。

【図 5】 本発明の実施形態に係るコンピュータ 900 のハードウェア構成の一例を示す。

【符号の説明】

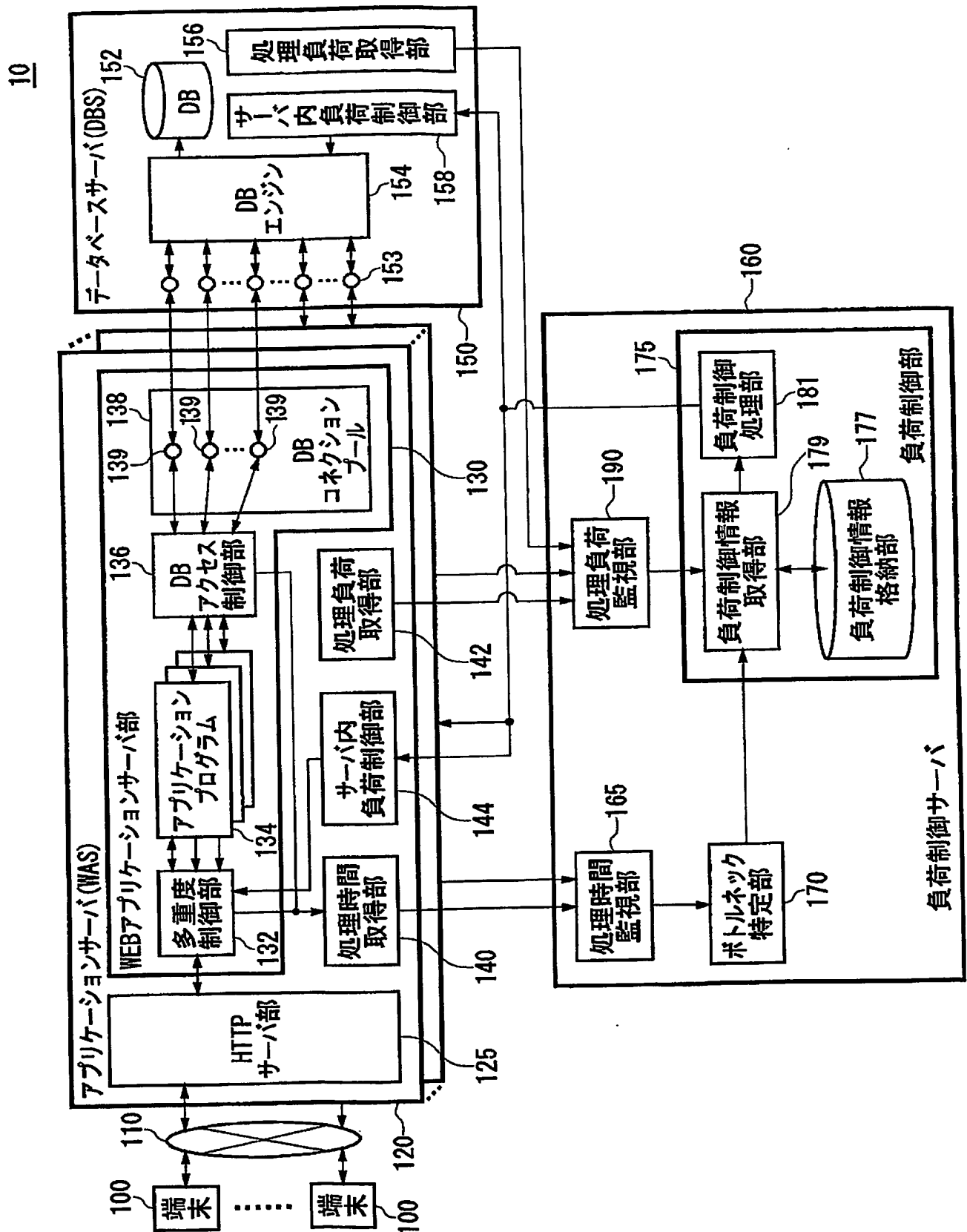
【0087】

- 10 情報システム
- 100 端末
- 110 ネットワーク
- 120 WAS

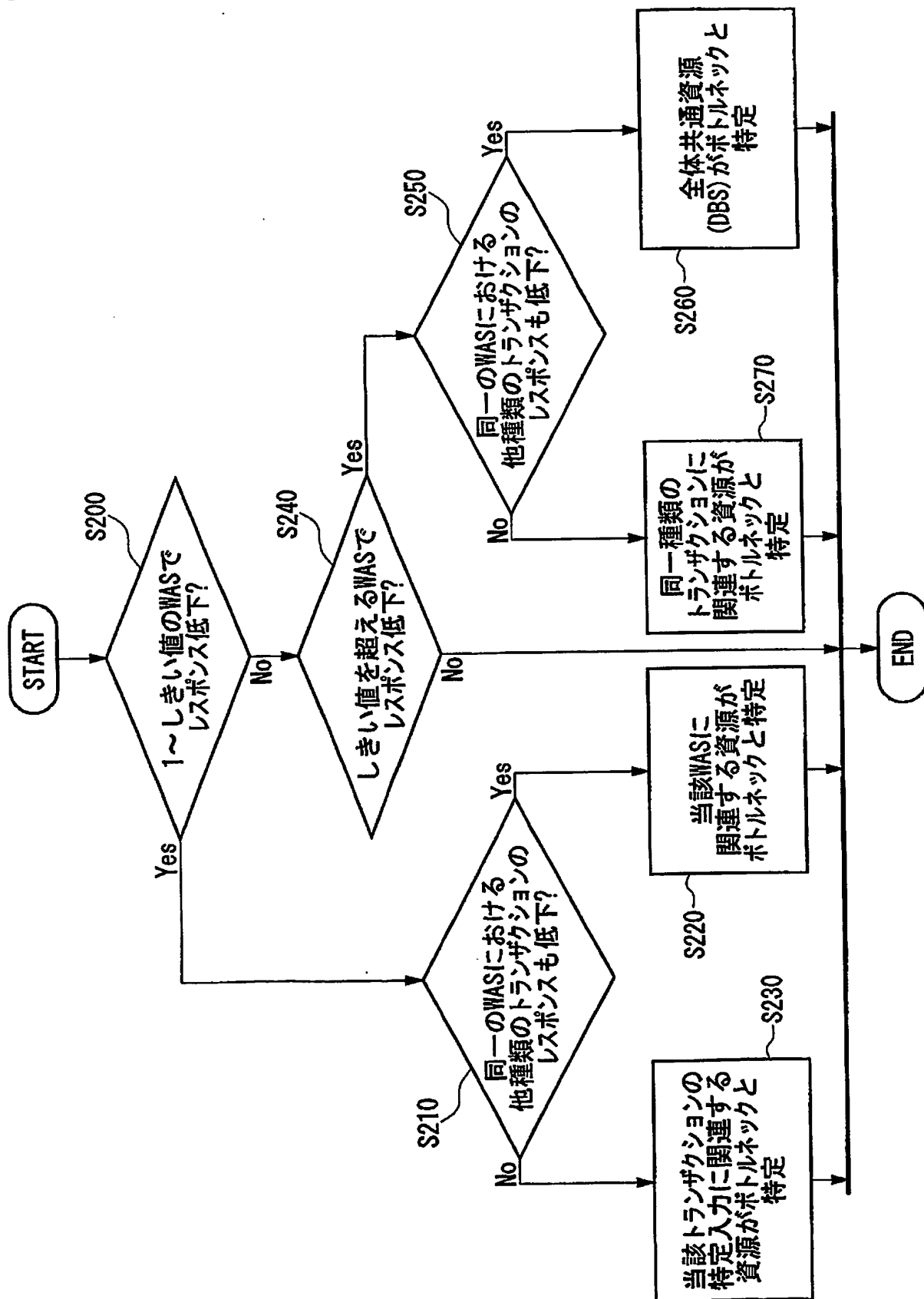
1 2 5 HTTPサーバ部
1 3 0 WEBアプリケーションサーバ部
1 3 2 多重度制御部
1 3 4 アプリケーションプログラム
1 3 6 DBアクセス制御部
1 3 8 DBコネクションプール
1 3 9 DBコネクション
1 4 0 処理時間取得部
1 4 2 処理負荷取得部
1 4 4 サーバ内負荷制御部
1 5 0 DBS
1 5 2 DB
1 5 3 DB処理要求 I / F 部
1 5 4 DBエンジン
1 5 6 処理負荷取得部
1 5 8 サーバ内負荷制御部
1 6 0 負荷制御サーバ
1 6 5 処理時間監視部
1 7 0 ボトルネック特定部
1 7 5 負荷制御部
1 7 7 負荷制御情報格納部
1 7 9 負荷制御情報取得部
1 8 1 負荷制御処理部
1 9 0 処理負荷監視部
9 0 0 コンピュータ
1 0 0 0 CPU
1 0 1 0 ROM
1 0 2 0 RAM
1 0 3 0 通信インターフェイス
1 0 4 0 記憶装置
1 0 5 0 フレキシブルディスク・ドライブ
1 0 6 0 CD-ROMドライブ
1 0 7 0 入出力チップ
1 0 7 5 グラフィック・コントローラ
1 0 8 0 表示装置
1 0 8 2 ホスト・コントローラ
1 0 8 4 入出力コントローラ
1 0 9 0 フレキシブルディスク
1 0 9 5 CD-ROM

【書類名】 図面

【図 1】



【図 2】



【図3】

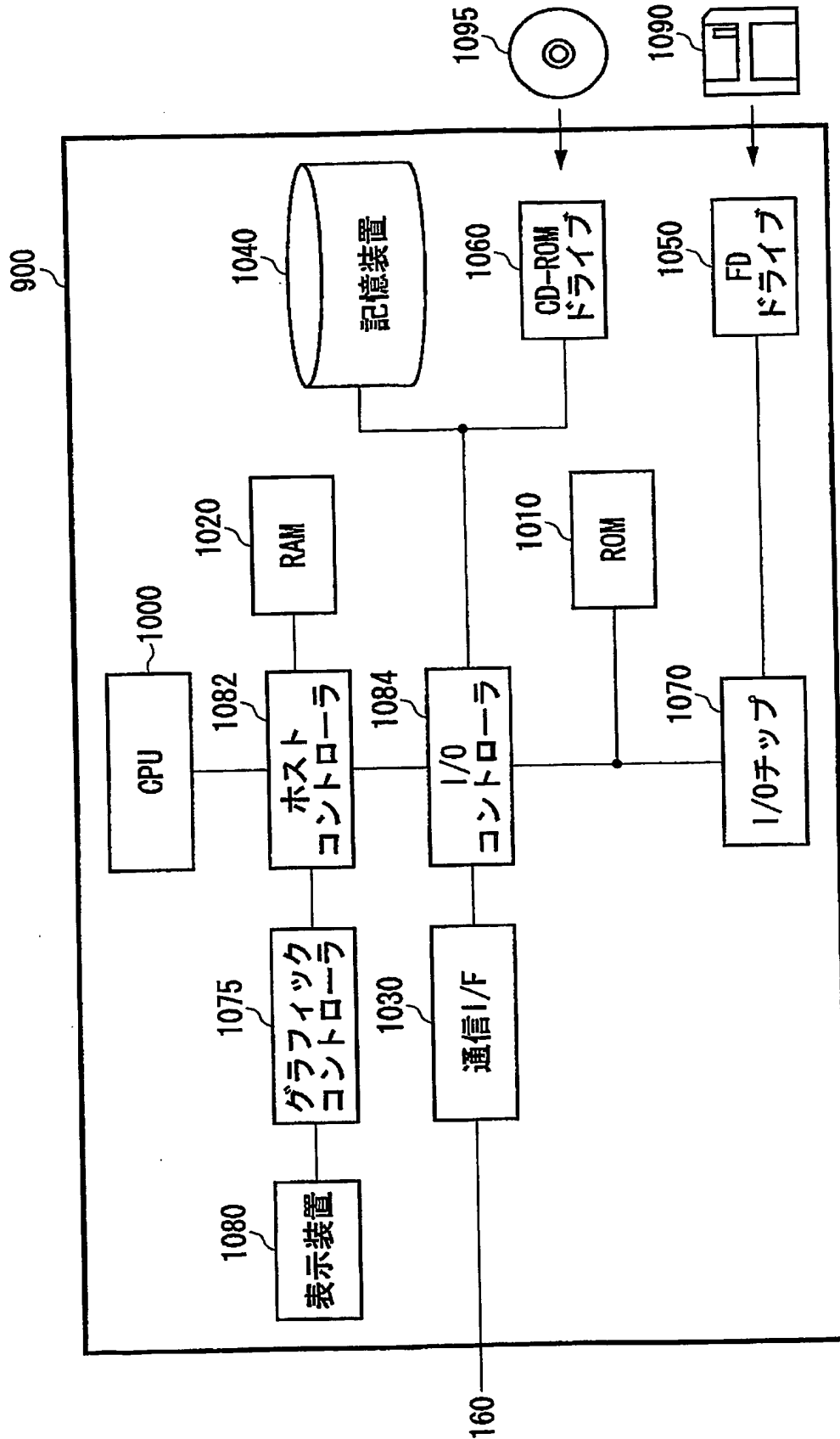
ボトルネック箇所の 分類	資源	検出条件	判定条件	解除条件
全体共通資源 (DBS)	プロセッサ	使用率が n 秒間平均 $x\%$ 以上	m 回続く	f 回検出されない
	ディスク (DB, LOG, 他)	使用率が n 秒間平均 $x\%$ 以上	m 回続く	f 回検出されない
	メモリ	ページング回数が n 秒平均 x 回以上	m 回続く	f 回検出されない
	DBバッファ	バッファミス率が n 秒平均 $x\%$ 以上	m 回続く	f 回検出されない
WAS	プロセッサ	使用率が n 秒平均 $x\%$ 以上	m 回続く	f 回検出されない
	ディスク	使用率が n 秒平均 $x\%$ 以上	m 回続く	f 回検出されない
	メモリ	ページング回数が n 秒平均 x 回以上	m 回続く	f 回検出されない
	JVMメモリ	GCが n 秒間に x 回以上	m 回続く	f 回検出されない
トランザクションに 関連する資源	DBコネクション	多重度が上限に達した	発生時	WAS停止時
	WASのプロセッサ	DBS処理時間/処理時間が $x\%$ 以下	発生時	n 秒間発生しない
	DBSのプロセッサ	DBS処理時間/処理時間が $x\%$ 以上 n 秒以内にDBSのプロセッサがボトルネックと判定	発生時	n 秒間発生しない
	DBSのディスク	DBS処理時間/処理時間が $x\%$ 以上 n 秒以内にDBSのディスクがボトルネックと判定	発生時	n 秒間発生しない

【図4】

177

ボトルネック	資源	負荷制御情報
全体共通資源 (DBS)	プロセッサ	負荷制御処理 3 → 1 → 4 → 2 → 5 → 6
	ディスク (DB)	負荷制御処理 2 → 1 → 6 → 5
	ディスク (LOG)	負荷制御処理 1 → 2 → 5 → 6
	ディスク (他)	負荷制御処理 1 → 5
	メモリ	負荷制御処理 1 → 5
	DBバッファ	負荷制御処理 1 → 5
WAS	プロセッサ	負荷制御処理 3 → 1 → 5
	ディスク	負荷制御処理 1 → 5
	メモリ	負荷制御処理 1 → 5
	JVMメモリ	負荷制御処理 1 → 5
	DBコネクション	負荷制御処理 1 → 5
	WASのプロセッサ	負荷制御処理 5
トランザクションに 関連する資源	DBSのプロセッサ	負荷制御処理 5 → 6
	DBSのディスク	負荷制御処理 5 → 6

【図 5】



【書類名】 要約書**【要約】**

【課題】 複数のアプリケーションサーバ及びデータベースサーバを備える情報システムにおいて、各アプリケーションサーバによるトランザクションの処理時間に応じて適切な負荷制御を行う。

【解決手段】 複数のアプリケーションサーバと、データベースサーバとを備える情報システムであって、アプリケーションサーバ毎に、当該アプリケーションサーバが受信したトランザクションをアプリケーションプログラムが処理する処理時間を監視する処理時間監視部と、アプリケーションサーバ毎の処理時間の監視結果に基づいて、処理時間が予め定められた許容範囲内でない前記アプリケーションサーバにボトルネックがあるかを特定するボトルネック特定部と、ボトルネックの原因があると特定されたアプリケーションサーバ上でアプリケーションプログラムを実行する多重度を低下させる負荷制御部とを備える情報システムを提供する。

【選択図】 図1

認定・付加情報

特許出願の番号	特願 2003-368773
受付番号	50301793030
書類名	特許願
担当官	伊藤 雅美 2132
作成日	平成 15 年 10 月 30 日

<認定情報・付加情報>

【特許出願人】

【識別番号】	592073101
【住所又は居所】	東京都港区六本木 3 丁目 2 番 12 号
【氏名又は名称】	日本アイ・ビー・エム株式会社

【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

【代理人】

【識別番号】	100108501
【住所又は居所】	神奈川県大和市下鶴間 1623 番 14 日本アイ・ビー・エム株式会社 知的所有権
【氏名又は名称】	上野 剛史

【復代理人】

【識別番号】	100104156
【住所又は居所】	東京都新宿区新宿 1 丁目 24 番 12 号 東信ビル 6 階 龍華国際特許事務所
【氏名又は名称】	龍華 明裕

【代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

特願 2 0 0 3 - 3 6 8 7 7 3

出 願 人 履 歴 情 報

識別番号

[5 9 2 0 7 3 1 0 1]

1. 変更年月日

1 9 9 2 年 4 月 3 日

[変更理由]

新規登録

住 所

東京都港区六本木 3 丁目 2 番 1 2 号

氏 名

日本アイ・ビー・エム株式会社